

# CHARACTERIZATION OF WORST-CASE GMRES

VANCE FABER<sup>\*</sup>, JÖRG LIESEN<sup>†</sup> AND PETR TICHÝ<sup>‡</sup>

**Abstract.** Given a matrix  $A$  and iteration step  $k$ , we study a best possible attainable upper bound on the GMRES residual norm that does not depend on the initial vector  $b$ . This quantity is called the worst-case GMRES approximation. We show that the worst case behavior of GMRES for the matrices  $A$  and  $A^T$  is the same, and we analyze properties of initial vectors for which the worst-case residual norm is attained. In particular, we show that such vectors satisfy a certain “cross equality”, and we characterize them as right singular vectors of the corresponding GMRES residual matrix. We show that the worst-case GMRES polynomial may not be uniquely determined, and we consider the relation between the worst-case and the ideal GMRES approximations, giving new examples in which the inequality between the two quantities is sharp at all iteration steps  $k \geq 3$ . Finally, we give a complete characterization of how the values of the approximation problems in the context of worst-case and ideal GMRES for a real matrix change, when one considers complex (rather than real) polynomials and initial vectors in these problems.

**Key words.** GMRES convergence, matrix approximation problems, minmax

**AMS subject classifications.** 65F10, 49K35, 41A52

**1. Introduction.** Let a nonsingular matrix  $A \in \mathbb{R}^{n \times n}$  and a vector  $b \in \mathbb{R}^n$  be given. Consider solving the system of linear algebraic equations  $Ax = b$  with the initial guess  $x_0 = 0$  using the GMRES method [11]. This method generates a sequence of iterates  $x_k \in \mathcal{K}_k(A, b) \equiv \text{span}\{b, Ab, \dots, A^{k-1}b\}$ ,  $k = 1, 2, \dots$ , so that the corresponding  $k$ th residual  $r_k \equiv b - Ax_k$  satisfies

$$(1.1) \quad \|r_k\| = \min_{p \in \pi_k} \|p(A)b\|.$$

Here  $\|\cdot\|$  denotes the Euclidean norm, and  $\pi_k$  denotes the set of real polynomials of degree at most  $k$  and with value one at the origin. Note that for a real matrix  $A$  and a real right hand side  $b$  the minimum in (1.1) is achieved for a real polynomial. Considering only real polynomials therefore does not represent any restriction.

It is clear from (1.1), that the sequence of GMRES residual norms  $\|r_k\|$ ,  $k = 1, 2, \dots$ , is nonincreasing. It terminates with  $r_k = 0$  if and only if  $k$  is equal to  $d(A, b)$ , the degree of the minimal polynomial of the vector  $b$  with respect to  $A$ . For each  $b$  we have  $d(A, b) \leq d(A)$ , the degree of the minimal polynomial of  $A$ .

A geometric characterization of the iterate  $x_k \in \mathcal{K}_k(A, b)$ , which is mathematically equivalent to (1.1), is given by

$$(1.2) \quad r_k \perp A\mathcal{K}_k(A, b).$$

To emphasize the dependence of the  $k$ th GMRES residual  $r_k$  on the given data  $A, b$  and  $k$  we will sometimes write

$$r_k = \text{GMRES}(A, b, k) \quad \text{or} \quad r_k = p_k(A)b,$$

<sup>\*</sup>Vanco Research, Big Pine Key, FL 33043 (vance.faber@gmail.com).

<sup>†</sup>Institute of Mathematics, Technical University of Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany (liesen@math.tu-berlin.de). The work of this author was supported by the Heisenberg Program of the Deutsche Forschungsgemeinschaft (DFG).

<sup>‡</sup>Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 2, 18207 Prague, Czech Republic (tichy@cs.cas.cz). This work was supported by the Grant Agency of the Czech Republic under grant No. P201/13-06684 S, and by the project M100301201 of the institutional support of the Academy of Sciences of the Czech Republic.

where  $p_k \in \pi_k$  is the  $k$ th GMRES polynomial of  $A$  and  $b$ , i.e., the polynomial that solves the minimization problem on the right hand side of (1.1). As long as  $r_k \neq 0$ , this polynomial is uniquely determined. The matrix  $p_k(A)$  is called the  $k$ th GMRES residual matrix of  $A$  and  $b$ . For further basic properties and algorithmic details of the GMRES method we refer to the original paper [11] or the books [2, 8, 10].

In the following we will assume without loss of generality that  $\|b\| = 1$ . A common approach for investigating the GMRES convergence behavior is to bound (1.1) independently of  $b$ . For each iteration step  $k$  the best possible bound on the GMRES residual norm that is independent of  $b$  is given by maximizing the right hand side of (1.1) over all unit norm vectors, i.e.,

$$(1.3) \quad \|r_k\| = \min_{p \in \pi_k} \|p(A)b\| \leq \max_{\|v\|=1} \min_{p \in \pi_k} \|p(A)v\| \equiv \Psi_k(A).$$

The quantity  $\Psi_k(A)$  is called the  $k$ th worst-case GMRES approximation. It is easy to see that the bound (1.3) is sharp in the sense that for each given  $A$  and  $k$  there exists a unit norm vector  $b$  so that the corresponding  $k$ th GMRES residual vector satisfies  $\|r_k\| = \Psi_k(A)$ . We will call such a vector  $b$ , the corresponding  $k$ th GMRES polynomial  $p_k$  and the corresponding  $k$ th GMRES residual matrix  $p_k(A)$  the  $k$ th worst-case GMRES initial vector, polynomial and residual matrix, respectively. If  $A$  is singular, then  $\Psi_k(A) = 1$  for all  $k \geq 0$  (to see this, simply take  $b$  as a unit norm vector in the kernel of  $A$ ). Hence only the case of a nonsingular matrix  $A$  is of interest in this context. For such  $A$  we have

$$1 \geq \Psi_1(A) \geq \dots \geq \Psi_{d(A)-1}(A) > \Psi_{d(A)}(A) = 0,$$

and therefore we only need to consider  $1 \leq k \leq d(A) - 1$ .

It is known that  $\Psi_k(A)$  for a fixed  $k$  is a continuous function on the open set of nonsingular matrices; see [5, Theorem 3.1] or [1, Theorem 2.5]. Moreover, it was shown in [1, Theorem 2.7] that  $\Psi_k(A) = 1$  for a nonsingular matrix  $A$ , if and only if zero is contained in some generalized field of values derived from the powers  $I, A, \dots, A^k$ . Most of the other previously published results on worst-case GMRES are devoted to studying the tightness of the inequality

$$(1.4) \quad \Psi_k(A) \leq \min_{p \in \pi_k} \|p(A)\| \equiv \varphi_k(A),$$

which is easily derived from (1.3) using the submultiplicativity property of the Euclidean norm. The quantity  $\varphi_k(A)$  is called the  $k$ th ideal GMRES approximation [4]. The polynomial for which the minimum is attained in (1.4) is called the  $k$ th ideal GMRES polynomial of  $A$ . This polynomial is uniquely determined; see [4, 9]. It was shown that (1.4) is an equality for normal matrices  $A$  and all  $k \geq 0$ , and for  $k = 1$  and any nonsingular  $A$  [3, 6]. Some nonnormal matrices  $A$  are known for which  $\Psi_k(A) < \varphi_k(A)$ , even  $\Psi_k(A) \ll \varphi_k(A)$ , for certain  $k$ ; see [1, 13].

The ideal GMRES approximation problem can be formulated as a semidefinite program (see [14]) and hence can be solved efficiently by standard software. On the other hand, we are unaware of any efficient algorithm for solving the worst-case GMRES approximation problem, so that in practice one needs to resort to a “general purpose” nonlinear solver to compute worst-case GMRES data. The difficult nonlinear nature of the worst-case GMRES approximation problem may be one of the reasons why this problem is less studied (both theoretically and numerically) than the ideal GMRES approximation problem.

This paper is mainly devoted to characterizations of the worst-case GMRES problem (1.3). We first show in Section 2 that  $\Psi_k(A) = \Psi_k(A^T)$ , and that worst-case initial vectors satisfy a certain “cross equality”. Next, in Section 3, we look at the worst-case GMRES approximation problem from the optimization point of view and show that  $k$ th worst-case GMRES initial vectors are always right singular vectors of the corresponding  $k$ th GMRES residual matrix. In Section 4 we prove that a  $k$ th worst-case GMRES polynomial may not be uniquely determined (unlike the  $k$ th ideal GMRES polynomial), and we give a numerical example for two different polynomials and corresponding initial vectors that both attain the same worst-case GMRES value at the same step  $k$ . In Section 5 we further study differences between the worst-case and the ideal GMRES approximations. In particular, we state a parameterized set of matrices  $A$  of arbitrary size  $2n$  (with  $n \geq 2$ ) for which the inequality in (1.4) is sharp for all  $k = 3, \dots, 2n - 1$ . In the previously published examples in [1, 13], a small matrix  $A$  is constructed for which the sharp inequality occurs for exactly one  $k$ . Finally, in Section 6 we analyze whether the values of the max-min approximation (1.3) and the min-max approximation (1.4) for a real matrix change if we consider the maximization over complex vectors and/or the minimization over complex polynomials. This analysis gives another indication for the difference between the two approximation problems.

**2. The cross equality.** In this section we generalize two results of Zavorin [15]. The first shows that  $\Psi_k(A) = \Psi_k(A^T)$  and the second concerns a special property of worst-case initial vectors (they satisfy the so-called “cross equality”). Note that Zavorin proved these results only for diagonalizable matrices using quite a complicated technique based on the decomposition of the corresponding Krylov matrix. Using a simple algebraic technique we prove these results for general matrices. All results presented in this section can easily be generalized from real to complex matrices.

**THEOREM 2.1.** *If  $A \in \mathbb{R}^{n \times n}$  is a nonsingular matrix, then  $\Psi_k(A) = \Psi_k(A^T)$  for all  $k = 1, \dots, d(A) - 1$ .*

*Proof.* Let  $1 \leq k \leq d(A) - 1$  and consider any unit norm vector  $b$  such that the corresponding  $k$ th GMRES residual vector  $r_k = p_k(A)b$  is nonzero. The defining property (1.2) of  $r_k$  means that  $\langle A^j b, r_k \rangle = 0$  for  $j = 1, \dots, k$ . Hence, for any  $q \in \pi_k$ ,

$$(2.1) \quad \|r_k\|^2 = \langle p_k(A)b, r_k \rangle = \langle b, r_k \rangle = \langle q(A)b, r_k \rangle = \langle b, q(A^T)r_k \rangle \leq \|q(A^T)r_k\|,$$

where the last inequality follows from the Cauchy-Schwarz inequality and  $\|b\| = 1$ .

If  $b$  is a unit norm  $k$ th worst-case GMRES initial vector and  $r_k$  is the corresponding  $k$ th GMRES residual vector, then the previous inequality means that

$$(2.2) \quad \|r_k\|^2 = \Psi_k^2(A) \leq \|q(A^T)r_k\|,$$

where  $q \in \pi_k$  is arbitrary. Dividing by  $\|r_k\|$  and taking the minimum over all  $q \in \pi_k$  we get

$$(2.3) \quad \Psi_k(A) \leq \min_{q \in \pi_k} \left\| q(A^T) \frac{r_k}{\|r_k\|} \right\| \leq \Psi_k(A^T).$$

Now we can reverse the roles of  $A$  and  $A^T$ , and then repeat the whole argument to obtain the opposite inequality, i.e.,  $\Psi_k(A^T) \leq \Psi_k(A)$ .  $\square$

The following theorem describes a special property of worst-case initial vectors: If we apply GMRES to  $A$  and a unit norm  $k$ th worst-case initial vector  $b$  giving at

step  $k$  the residual vector  $r_k$ , and then  $k$  steps of GMRES to  $A^T$  and the initial vector  $r_k/\|r_k\|$ , we obtain again the original initial vector  $b$  (up to a scaling factor).

**THEOREM 2.2.** *Let  $A \in \mathbb{R}^{n \times n}$  be a nonsingular matrix, and let  $1 \leq k \leq d(A) - 1$ . If  $b \in \mathbb{R}^n$  is a unit norm  $k$ th worst-case GMRES initial vector and*

$$r_k \equiv \text{GMRES}(A, b, k), \quad s_k \equiv \text{GMRES}\left(A^T, \frac{r_k}{\|r_k\|}, k\right),$$

then

$$\|s_k\| = \|r_k\| = \Psi_k(A) \quad \text{and} \quad b = \frac{s_k}{\Psi_k(A)}.$$

*Proof.* Let  $b$  be a unit norm  $k$ th worst-case GMRES initial vector and let  $r_k = \text{GMRES}(A, b, k)$ . In addition, let  $s_k = \text{GMRES}(A^T, r_k/\|r_k\|, k)$  and let  $q_k$  be the corresponding  $k$ th GMRES polynomial. Using this polynomial in (2.2) yields

$$\|r_k\| = \Psi_k(A) \leq \left\| q_k(A^T) \frac{r_k}{\|r_k\|} \right\| = \|s_k\| \leq \Psi_k(A^T).$$

However, as shown in Theorem 2.1, equality holds throughout, which shows the first assertion.

Moreover, since  $\|r_k\| = \|s_k\|$ , the (Cauchy-Schwarz) inequality on the right of (2.1) is an equality for the given  $b$  and  $q = q_k$ , i.e.,

$$\langle b, q_k(A^T)r_k \rangle = \|q_k(A^T)r_k\|.$$

Since  $\|b\| = 1$ , this happens if and only if

$$b = \frac{q_k(A^T)r_k}{\|q_k(A^T)r_k\|} = \frac{q_k(A^T)r_k}{\|r_k\|\|r_k\|} = \frac{s_k}{\|r_k\|},$$

which finishes the proof.  $\square$

The previous theorem shows that if  $b$  is a unit norm  $k$ th worst-case GMRES initial vector, then (with the same notation as in the proof above)

$$\Psi_k(A)b = s_k = q_k(A^T) \frac{r_k}{\|r_k\|} = q_k(A^T)p_k(A) \frac{b}{\|r_k\|},$$

or, equivalently,

$$(2.4) \quad q_k(A^T)p_k(A)b = \Psi_k^2(A)b.$$

In other words,  $b$  is an eigenvector of the matrix  $q_k(A^T)p_k(A)$  with the corresponding eigenvalue  $\Psi_k^2(A)$ . In Corollary 3.7 we will show that  $q_k = p_k$ , i.e., that  $b$  is a right singular vector of the  $k$ th worst-case GMRES residual matrix  $p_k(A)$ .

To further investigate vectors with the special property introduced in Theorem 2.2 we use the following definition.

**DEFINITION 2.3.** *Let  $A \in \mathbb{R}^{n \times n}$  be nonsingular. We say that a unit norm vector  $b \in \mathbb{R}^n$  with  $d(A, b) > k$  satisfies the cross equality for  $A$  and the step  $k \geq 1$ , if*

$$b = \frac{s_k}{\|s_k\|}, \quad \text{where} \quad s_k \equiv \text{GMRES}\left(A^T, \frac{r_k}{\|r_k\|}, k\right), \quad r_k \equiv \text{GMRES}(A, b, k).$$

---

**Algorithm 1** (Cross iterations 1)

---

```
 $b^{(0)} = b,$   
for  $j = 1, 2, \dots$  do  
   $r_k^{(j)} = \text{GMRES}(A, b^{(j-1)}, k)$   
   $c^{(j-1)} = r_k^{(j)} / \|r_k^{(j)}\|$   
   $s_k^{(j)} = \text{GMRES}(A^T, c^{(j-1)}, k)$   
   $b^{(j)} = s_k^{(j)} / \|s_k^{(j)}\|$   
end for
```

---

Inspired by Theorem 2.2 we define the iterative process shown in Algorithm 1. To analyze this algorithm, let us denote

$$r_k^{(j)} = p_k^{(j)}(A)b^{(j-1)} \quad \text{and} \quad s_k^{(j)} = q_k^{(j)}(A^T)c^{(j-1)}.$$

Using  $q = q_k^{(j)}$  in (2.1) we then get

$$\|r_k^{(j)}\|^2 \leq \|q_k^{(j)}(A^T)r_k^{(j)}\| = \|r_k^{(j)}\| \|q_k^{(j)}(A^T)c^{(j-1)}\| = \|r_k^{(j)}\| \|s_k^{(j)}\|.$$

Now consider (2.1) with the roles of  $A$  and  $A^T$  reversed, i.e.,

$$\begin{aligned} \|s_k^{(j)}\|^2 &= \langle q_k^{(j)}(A^T)c^{(j-1)}, s_k^{(j)} \rangle = \langle c^{(j-1)}, q(A)s_k^{(j)} \rangle = \|s_k^{(j)}\| \langle c^{(j-1)}, q(A)b^{(j)} \rangle \\ &\leq \|s_k^{(j)}\| \|q(A)b^{(j)}\|, \end{aligned}$$

for all  $q \in \pi_k$ . We can choose  $q = p_k^{(j+1)}$  and thus obtain  $\|s_k^{(j)}\| \leq \|r_k^{(j+1)}\|$ . In summary, we have shown that

$$(2.5) \quad \|r_k^{(j)}\| \leq \|s_k^{(j)}\| \leq \|r_k^{(j+1)}\| \leq \|s_k^{(j+1)}\| \leq \Psi_k(A), \quad j = 1, 2, \dots$$

Hence the sequences of norms  $\|r_k^{(j)}\|$  and  $\|s_k^{(j)}\|$ ,  $j = 1, 2, \dots$ , interlace each other, are both nondecreasing, and are both bounded by  $\Psi_k(A)$ . This implies that both sequences converge to the same limit, which does not exceed  $\Psi_k(A)$ .

Consequently, for any initial vector  $b^{(0)}$ , Algorithm 1 converges to a vector that satisfies the cross equality for  $A$  and step  $k$ . If  $b^{(0)}$  satisfies the cross equality for  $A$  and step  $k$ , then trivially equality holds in (2.5) for all  $j$ . On the other hand, if equality holds in (2.5) for one  $j$ , then, using (2.1),

$$\langle b^{(j)}, q_k^{(j)}(A^T)r_k^{(j)} \rangle = \|r_k^{(j)}\|^2 = \|q_k^{(j)}(A^T)r_k^{(j)}\| = \|s_k^{(j)}\|,$$

and we have reached a vector that satisfies the cross equality.

From the above it is clear that the cross equality represents a necessary condition for a vector  $b^{(0)}$  to be a worst-case initial vector. On the other hand, we can ask whether this condition is sufficient, or, at least, whether the vectors that satisfy the cross equality are in some sense special. To investigate this question we present the following lemma.

**LEMMA 2.4.** *Let  $A \in \mathbb{R}^{n \times n}$  be nonsingular,  $k \geq 1$ , and  $b \in \mathbb{R}^n$  be a unit norm initial vector with  $d(A, b) > k$ . If  $r_k = \text{GMRES}(A, b, k)$ , then  $d(A^T, r_k) > k$ , and  $b$*

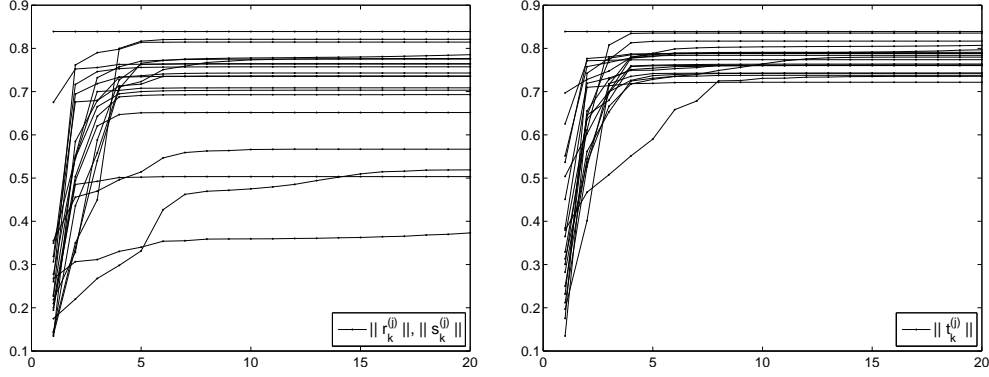


FIG. 2.1. Cross iterations for random initial vectors.

satisfies the cross for  $A$  and the step  $k$  if and only if  $b \in \mathcal{K}_{k+1}(A^T, r_k)$ . In particular, each unit norm vector  $b$  with  $d(A, b) = n$  satisfies the cross equality for  $A$  and the step  $k = n - 1$ .

*Proof.* The nonzero GMRES residual  $r_k \in b + A\mathcal{K}_k(A, b) \subset \mathcal{K}_{k+1}(A, b)$  is uniquely determined by the orthogonality conditions (1.2), which can be written as

$$0 = \langle A^j b, r_k \rangle = \langle b, (A^T)^j r_k \rangle, \quad \text{for } j = 1, \dots, k,$$

or, equivalently,

$$(2.6) \quad b \perp A^T \mathcal{K}_k(A^T, r_k).$$

Now let  $s_k \equiv \text{GMRES}(A^T, r_k/\|r_k\|, k)$ . From (2.1) we know that  $\|s_k\| \geq \|r_k\| > 0$ , i.e.  $d(A^T, r_k) > k$ , and

$$(2.7) \quad s_k \in \frac{r_k}{\|r_k\|} + A^T \mathcal{K}_k(A^T, r_k) \subset \mathcal{K}_{k+1}(A^T, r_k), \quad s_k \perp A^T \mathcal{K}_k(A^T, r_k).$$

If  $b$  satisfies the cross equality for  $A$  and the step  $k$ , then  $b = s_k/\|s_k\|$  and (2.7) implies that  $b \in \mathcal{K}_{k+1}(A^T, r_k)$ . On the other hand, if  $b \in \mathcal{K}_{k+1}(A^T, r_k)$ , then  $\langle b, r_k \rangle = \|r_k\|^2 \neq 0$  and (2.6) imply that  $b = s_k/\|s_k\|$ .

For  $k = n - 1$ , we have  $\mathcal{K}_{k+1}(A^T, r_k) = \mathbb{R}^n$ , i.e.  $b \in \mathcal{K}_{k+1}(A^T, r_k)$  is always satisfied.  $\square$

To give a numerical example for Algorithm 1 we consider  $A$  being the Jordan block  $J_\lambda$  of size 11 with the eigenvalue  $\lambda = 1$ , and we choose  $k = 5$ . In this case, the ideal GMRES matrix  $\varphi_5(A)$  has a simple maximal singular value, as numerically observed in [12]. Using the results of Greenbaum and Gurvits in [3] we know that then  $\Psi_5(J_\lambda) = \varphi_5(J_\lambda)$ , and, moreover, that the corresponding worst-case initial vector is the right singular vector that corresponds to the maximal singular value of the ideal GMRES matrix  $\varphi_5(A)$ . Hence, in this case the 5th worst-case initial vector is uniquely determined up to scaling.

In the left part of Fig. 2.1 we show the results of Algorithm 1 started with 20 random unit norm initial vectors. Each line represents the sequence  $\|r_k^{(j)}\|, \|s_k^{(j)}\|$ , for  $j = 1, \dots, 10$ . In the end of each of the 20 runs we get a vector that satisfies (up to a small inaccuracy) the cross equality for  $J_\lambda$  and  $k = 5$ . We can observe that there are

many initial vectors that satisfy the cross equality, and there seems to be no special structure in the norms that are attained in the end. In particular, none of the 20 runs results in a 5th worst-case initial vector for which the norm  $\Psi_5(A)$  is attained (this value is visualized by the highest horizontal line in the figure).

We will now slightly modify the cross iteration Algorithm 1. Having a initial vector  $b^{(j-1)}$  we always apply both, GMRES with  $A$  as well as GMRES with  $A^T$ , and look at the resulting GMRES residual norm. We take as a resulting residual the one with the greater norm; see Algorithm 2. After the process converges, we get again a vector that satisfies the cross equality.

---

**Algorithm 2** (Cross iterations 2)

---

```

 $b^{(0)} = b,$ 
for  $j = 1, 2, \dots$  do
   $v = \text{GMRES}(A, b^{(j-1)}, k)$ 
   $w = \text{GMRES}(A^T, b^{(j-1)}, k)$ 
  if  $\|v\| < \|w\|$  then
     $t_k^{(j)} = w$ 
  else
     $t_k^{(j)} = v$ 
  end if
   $b^{(j)} = t_k^{(j)} / \|t_k^{(j)}\|$ 
end for

```

---

This strategy is a little better than the original one when looking for a worst-case initial vector; see Fig. 2.1. While it is usually not sufficient to find a worst-case vector, one at least can find a reasonable initial point for an optimization procedure that solves the nonlinear worst-case GMRES approximation problem.

**3. Optimization point of view.** Let a nonsingular matrix  $A \in \mathbb{R}^{n \times n}$  and a positive integer  $k < d(A)$  be given. For vectors  $c = [c_1, \dots, c_k]^T \in \mathbb{R}^k$  and  $v \in \mathbb{R}^n$ , we define the function

$$(3.1) \quad f(c, v) \equiv \|p(A; c)v\|^2 = v^T p(A; c)^T p(A; c)v,$$

where

$$p(z; c) = 1 - \sum_{j=1}^k c_j z^j.$$

Equivalently, we can express the function  $f(c, v)$  using the matrix

$$K(v) \equiv [Av, A^2v, \dots, A^kv]$$

as

$$(3.2) \quad f(c, v) = \|v - K(v)c\|^2 = v^T v - 2v^T K(v)c + c^T K(v)^T K(v)c.$$

(Here only the dependence on  $v$  is expressed in the notation  $K(v)$ , because  $A$  and  $k$  are both fixed.) Note that  $K(v)^T K(v)$  is the Gramian matrix of the vectors  $Av, A^2v, \dots, A^kv$ ,

$$K(v)^T K(v) = [v^T (A^T)^i A^j v]_{i,j=1,\dots,k}.$$

Next, we define the function

$$g(v) \equiv \min_{c \in \mathbb{R}^k} f(c, v),$$

which represents the  $k$ th squared GMRES residual norm for the matrix  $A$  and the initial vector  $v$ , and we denote

$$\Omega \equiv \{u \in \mathbb{R}^n : d(A, u) \geq k\}, \quad \Gamma \equiv \{u \in \mathbb{R}^n : d(A, u) < k\}.$$

The set  $\Gamma$  is a closed subset,  $\Omega$  is an open subset of  $\mathbb{R}^n$ , and  $\mathbb{R}^n = \Omega \cup \Gamma$ . Note that  $g(v) > 0$  for all  $v \in \Omega$  and  $g(v) = 0$  for all  $v \in \Gamma$ . The following lemma is a special case of [1, Proposition 2.2] for real data and nonsingular  $A$ .

LEMMA 3.1. *In the previous notation, the function  $g(v)$  is a continuous function of  $v \in \mathbb{R}^n$ , i.e.,  $g \in C^0(\mathbb{R}^n)$ , and it is an infinitely differentiable function of  $v \in \Omega$ , i.e.,  $g \in C^\infty(\Omega)$ . Moreover,  $\Gamma$  has measure zero in  $\mathbb{R}^n$ .*

We next characterize the minimizer of the function  $f(c, v)$  as a function of  $v$ .

LEMMA 3.2. *For each given  $v \in \Omega$ , the problem*

$$\min_{c \in \mathbb{R}^k} f(c, v)$$

*has the unique minimizer*

$$\gamma(v) = (K(v)^T K(v))^{-1} K(v)^T v \in \mathbb{R}^k.$$

*As a function of  $v \in \Omega$ , this minimizer satisfies  $\gamma(v) \in C^\infty(\Omega)$ . Given  $v \in \Omega$ ,  $(\gamma(v), v)$  is the only point in  $\mathbb{R}^k \times \Omega$  with*

$$\nabla_c f(\gamma(v), v) = 0.$$

*Proof.* Since  $v \in \Omega$  and  $A$  is nonsingular, the vectors  $Av, A^2v, \dots, A^kv$  are linearly independent and  $K(v)^T K(v)$  is symmetric and positive definite. Therefore, if  $v \in \Omega$  is fixed, (3.2) is a quadratic functional in  $c$ , which attains its unique global minimum at the stationary point

$$\gamma(v) = (K(v)^T K(v))^{-1} K(v)^T v.$$

The function  $\gamma(v)$  is a well defined rational function of  $v \in \Omega$ , and thus  $\gamma(v) \in C^\infty(\Omega)$ . Note that the vector  $\gamma(v)$  contains the coefficients of the  $k$ th GMRES polynomial that corresponds to the initial vector  $v \in \Omega$ .  $\square$

As stated in Lemma 3.1,  $g(v)$  is a continuous function on  $\mathbb{R}^n$ , and thus it is also continuous on the unit sphere

$$S \equiv \{u \in \mathbb{R}^n : \|u\| = 1\}.$$

Since  $S$  is a compact set and  $g(v)$  is continuous on this set, it attains its minimum and maximum on  $S$ .

We are interested in the characterization of points  $(\tilde{c}, \tilde{v}) \in \mathbb{R}^k \times S$  such that

$$(3.3) \quad f(\tilde{c}, \tilde{v}) = \max_{v \in S} \min_{c \in \mathbb{R}^k} f(c, v) = \max_{v \in S} g(v).$$



This is the worst-case GMRES problem (1.3). Since  $g(v) = 0$  for all  $v \in \Gamma$ , we have

$$\max_{v \in S} g(v) = \max_{v \in S \cap \Omega} g(v).$$

To characterize the points  $(\tilde{c}, \tilde{v}) \in \mathbb{R}^k \times S$  that satisfy (3.3), we define for every  $c \in \mathbb{R}^k$  and  $v \neq 0$  the two functions

$$F(c, v) \equiv f\left(c, \frac{v}{\|v\|}\right) = \frac{f(c, v)}{v^T v}, \quad G(v) \equiv g\left(\frac{v}{\|v\|}\right) = \frac{g(v)}{v^T v}.$$

Clearly, for any  $\alpha \neq 0$ , we have

$$F(c, \alpha v) = F(c, v), \quad G(\alpha v) = G(v).$$

LEMMA 3.3. *It holds that  $G(v) \in C^\infty(\Omega)$ . A vector  $\tilde{v} \in \Omega \cap S$  satisfies*

$$g(\tilde{v}) \geq g(v) \quad \text{for all } v \in S$$

*if and only if  $\tilde{v} \in \Omega \cap S$  satisfies*

$$G(\tilde{v}) \geq G(v) \quad \text{for all } v \in \mathbb{R}^n \setminus \{0\}.$$

*Proof.* Since  $g(v) \in C^\infty(\Omega)$  and  $0 \notin \Omega$ , it holds also  $G(v) \in C^\infty(\Omega)$ . If  $\tilde{v} \in \Omega \cap S$  is a maximum of  $G(v)$ , then  $\alpha \tilde{v}$  is a maximum as well, so the equivalence is obvious.  $\square$

THEOREM 3.4. *The vectors  $\tilde{c} \in \mathbb{R}^k$  and  $\tilde{v} \in S \cap \Omega$  that solve the problem*

$$\max_{v \in S} \min_{c \in \mathbb{R}^n} f(c, v)$$

*satisfy*

$$(3.4) \quad \nabla_c F(\tilde{c}, \tilde{v}) = 0, \quad \nabla_v F(\tilde{c}, \tilde{v}) = 0,$$

*i.e.,  $(\tilde{c}, \tilde{v})$  is a stationary point of the function  $F(c, v)$ .*

*Proof.* Obviously, for any  $v \in \Omega$ ,

$$F(\gamma(v), v) = \frac{f(\gamma(v), v)}{v^T v} \leq \frac{f(c, v)}{v^T v} = F(c, v) \quad \text{for all } c \in \mathbb{R}^k,$$

i.e.,  $\gamma(v)$  also minimizes the function  $F(c, v)$  and that

$$\nabla_c F(\gamma(v), v) = 0, \quad v \in \Omega.$$

We know that  $g(v)$  attains its maximum on  $S$  at some point  $\tilde{v} \in \Omega \cap S$ . Therefore,  $G(v)$  attains its maximum also at  $\tilde{v}$ . Since  $G(v) \in C^\infty(\Omega)$ , it has to hold that

$$\nabla G(\tilde{v}) = 0.$$

Denoting  $\tilde{c} = \gamma(\tilde{v})$  and writing the function  $G(v)$  as  $G(v) = F(\gamma(v), v)$  we get

$$(3.5) \quad \nabla G(\tilde{v}) = 0 = \nabla_v \gamma(\tilde{v}) \nabla_c F(\tilde{c}, \tilde{v}) + \nabla_v F(\tilde{c}, \tilde{v}),$$

where  $\nabla_v \gamma(\tilde{v})$  is the  $n \times k$  Jacobian matrix of the function  $\gamma(v) : \mathbb{R}^n \rightarrow \mathbb{R}^k$  at the point  $\tilde{v}$ . Here we used the standard chain rule for multivariate functions. Since  $\tilde{v} \in \Omega \cap S$ , we know from the previous that  $\nabla_c F(\tilde{c}, \tilde{v}) = 0$ , and, therefore, using (3.5),  $\nabla_v F(\tilde{c}, \tilde{v}) = 0$ .  $\square$

**THEOREM 3.5.** *If  $(\tilde{c}, \tilde{v})$  is a solution of the problem (3.3), then  $\tilde{v}$  is a right singular vector of the matrix  $p(A; \tilde{c})$ .*

*Proof.* Since  $(\tilde{c}, \tilde{v})$  solves the problem (3.3), we have  $0 = \nabla_v F(\tilde{c}, \tilde{v})$ . Writing  $F(c, v)$  as a Rayleigh quotient,

$$F(c, v) = \frac{v^T p(A; c)^T p(A, c) v}{v^T v},$$

we ask when  $\nabla_v F(c, v) = 0$ ; for more details see [7, pp. 114–115]. By differentiating  $F(c, v)$  with respect to  $v$  we get

$$0 = \frac{2p(A; c)^T p(A, c) v \|v\|^2 - 2v^T p(A; c)^T p(A, c) v v}{(v^T v)^2}$$

and the condition  $0 = \nabla_v F(\tilde{c}, \tilde{v})$  is equivalent to

$$p(A; \tilde{c})^T p(A, \tilde{c}) \tilde{v} = F(\tilde{c}, \tilde{v}) \tilde{v}.$$

In other words,  $\tilde{v}$  is a right singular vector of  $p(A; \tilde{c})$  and  $\sigma = \sqrt{F(\tilde{c}, \tilde{v})}$  is the corresponding singular value.  $\square$

**THEOREM 3.6.** *A point  $(\tilde{c}, \tilde{v}) \in \mathbb{R}^k \times S$  that solves the problem (3.3) is a stationary point of  $F(c, v)$  in which the maximal value of  $F(c, v)$  is attained.*

*Proof.* Using Theorem 3.4 we know that any solution  $(\tilde{c}, \tilde{v}) \in \mathbb{R}^k \times S$  of (3.3) is a stationary point of  $F(c, v)$ . On the other hand, if  $(\hat{c}, \hat{v}) \in \mathbb{R}^k \times S$  satisfies

$$\nabla_v F(\hat{c}, \hat{v}) = 0, \quad \nabla_c F(\hat{c}, \hat{v}) = 0,$$

then  $p(A; \hat{c})$  is the GMRES polynomial that corresponds to  $\hat{v}$  and

$$F(\hat{c}, \hat{v}) = \|p(A; \hat{c}) \hat{v}\|^2 \leq \|p(A; \tilde{c}) \tilde{v}\|^2 = F(\tilde{c}, \tilde{v}).$$

Hence,  $(\tilde{c}, \tilde{v})$  is a stationary point of  $F(c, v)$  in which the maximal value of  $F(c, v)$  is attained.  $\square$

As a consequence of previous results we can formulate the following corollary.

**COROLLARY 3.7.** *Let  $A \in \mathbb{R}^{n \times n}$  be a nonsingular matrix and let  $1 \leq k \leq d(A) - 1$ . Let  $b$  be a  $k$ th unit norm worst-case GMRES initial vector and let  $p_k \in \pi_k$  be the corresponding  $k$ th worst-case GMRES polynomial. Then  $p_k$  is also the  $k$ th worst-case GMRES polynomial for  $A^T$  and the initial vector  $r_k / \|r_k\|$ .*

*Proof.* Using Theorem 3.5 and Theorem 3.6 we know that

$$(3.6) \quad \Psi_k^2(A) b = p_k(A^T) p_k(A) b,$$

i.e., that  $b$  is a right singular vector of the GMRES residual matrix  $p_k(A)$  that corresponds to the maximal value of  $F(\tilde{c}, \tilde{v})$ , i.e., to  $\Psi_k^2(A)$ . From (2.4) we also know that

$$(3.7) \quad \Psi_k^2(A) b = q_k(A^T) p_k(A) b$$

where  $q_k$  is the GMRES polynomial that corresponds to  $A^T$  and the initial vector  $r_k$ . Comparing (3.6) and (3.7), and using the uniqueness of GMRES polynomials it follows that  $p_k = q_k$ .  $\square$

**4. Non-uniqueness of worst-case GMRES polynomials.** In this section we prove that a worst-case GMRES polynomial may not be uniquely determined, and we give a numerical example for the occurrence of a non-unique case. Our results are based on Toh's parameterized family of (nonsingular) matrices

$$(4.1) \quad A = A(\omega, \varepsilon) = \begin{bmatrix} 1 & \varepsilon & & \\ & -1 & \frac{\omega}{\varepsilon} & \\ & & 1 & \varepsilon \\ & & & -1 \end{bmatrix} \in \mathbb{R}^{4 \times 4}, \quad 0 < \omega < 2, \quad 0 < \varepsilon.$$

Toh used these matrices in [13] to show that  $\Psi_3(A)/\varphi_3(A) \rightarrow 0$  for  $\varepsilon \rightarrow 0$  and each  $\omega \in (0, 2)$  [13, Theorem 2.3]. In other words, he proved that the ratio of the worst-case and ideal GMRES approximations can be arbitrarily small.

**THEOREM 4.1.** *If  $p_k(z)$  is a  $k$ th worst-case GMRES polynomial of  $A$  in (4.1), then  $p_k(-z)$  is also a  $k$ th worst-case GMRES polynomial of  $A$ .*

*In particular,  $p_3(z) \neq p_3(-z)$ , so the third worst-case GMRES polynomial of  $A$  is not uniquely determined.*

*Proof.* Let  $b$  be any unit norm  $k$ th worst-case initial vector of  $A$ , and consider the orthogonal similarity transformation

$$A = -QA^TQ^T, \quad Q = \begin{bmatrix} & & & 1 \\ & & -1 & \\ & 1 & & \\ -1 & & & \end{bmatrix}.$$

Then

$$p_k(A)b = Qp_k(-A^T)Q^Tb \quad \text{and} \quad \Psi_k(A) = \|p_k(A)b\| = \|p_k(-A^T)w\| = \Psi_k(A^T),$$

where  $w = Q^Tb$ . In other words,  $p_k(-z)$  is a  $k$ th worst-case GMRES polynomial for  $A^T$  and, using Corollary 3.7, it is also a  $k$ th worst-case GMRES polynomial for the matrix  $A$ .

Let  $p_3(z) \in \pi_3$  be any third worst-case GMRES polynomial for the matrix  $A$ . To show that  $p_3(-z) \neq p_3(z)$  it suffices to show that  $p_3(z)$  contains odd powers of  $z$ , i.e., that

$$(4.2) \quad p_3(z) \neq 1 - \beta z^2 \quad \text{for any } \beta \in \mathbb{R}.$$

Define the matrix

$$B \equiv \begin{bmatrix} 1 & 0 & \omega & 0 \\ & 1 & 0 & \omega \\ & & 1 & 0 \\ & & & 1 \end{bmatrix} = A^2.$$

From [13, Theorem 2.1] we know that the (uniquely determined) third ideal GMRES polynomial of  $A$  is of the form

$$(4.3) \quad p_*(z) = 1 + (\alpha - 1)z^2, \quad \alpha = \frac{2\omega^2}{4 + \omega^2}.$$

Therefore,

$$\min_{p \in \pi_3} \|p(A)\| = \min_{p \in \pi_1} \max_{\|v\|=1} \|p(B)v\| = \max_{\|v\|=1} \min_{p \in \pi_1} \|p(B)v\|,$$

where the last equality follows from the fact that the ideal and worst-case GMRES approximations are equal for  $k = 1$  [6, 3]. If a third worst-case polynomial of  $A$  is of the form  $1 - \beta z^2$  for some  $\beta$ , then

$$\Psi_3(A) = \max_{\|v\|=1} \min_{p \in \pi_3} \|p(A)v\| = \max_{\|v\|=1} \min_{p \in \pi_1} \|p(B)v\| = \min_{p \in \pi_3} \|p(A)\| = \varphi_3(A).$$

This, however, contradicts the main result by Toh that  $\Psi_3(A) < \varphi_3(A)$ ; see [13, Theorem 2.2].  $\square$

To compute examples of worst-case GMRES polynomials for the Toh matrix (4.1) numerically we chose  $\varepsilon = 0.1$  and  $\omega = 1$ , and we used the function `fminsearch` from Matlab's Optimization Toolbox. We computed the value

$$\Psi_3(A) = 0.4579$$

(we present the numerical results only to 4 digits) with the corresponding third worst-case initial vector

$$b = [-0.6376, 0.0471, 0.2188, 0.7371]^T$$

and the worst-case GMRES polynomial

$$p_3(z) = -0.025z^3 - 0.895z^2 + 0.243z + 1 = \frac{-1}{39.9}(z - 1.181)(z + 0.939)(z + 35.96).$$

One can numerically check that  $b$  is the right singular vector of  $p_3(A)$  that corresponds to the second maximal singular value of  $p_3(A)$ . From Theorem 4.1 we know that  $q_3(z) \equiv p_3(-z)$  is also a third worst-case GMRES polynomial. One can now find the corresponding worst-case initial vector leading to the polynomial  $q_3$  using the singular value decomposition (SVD)

$$p_3(A) = USV^T,$$

where the singular values are ordered nonincreasingly on the diagonal of  $S$ . We know (by numerical observation) that  $b$  is the second column of  $V$ . We now compute the SVD of  $q_3(A)$ , and define the corresponding initial vector as the right singular vector that corresponds to the second maximal singular value of  $q_3(A)$ . It holds that

$$p_3(A^T) = p_3(A)^T = VSU^T.$$

Since  $A^T = -QAQ^T$ , we get  $Qp_3(-A)Q^T = VSU^T$ , or, equivalently,

$$q_3(A) = (Q^TV)S(Q^TU)^T.$$

So, the columns of the matrix  $Q^TU$  are right singular vectors of  $q_3(A)$  and the vector  $Q^Tu_2$ , where  $u_2$  is the second column of  $U$ , is the worst-case initial vector that gives the worst-case GMRES polynomial  $q_3(z) = p_3(-z)$ .

**5. Ideal versus worst-case GMRES phenomenon.** As mentioned above, Toh [13] as well as Faber, Joubert, Knill, and Manteuffel [1] have shown that worst-case GMRES and ideal GMRES are different approximation problems in the sense that there exist matrices  $A$  and iteration steps  $k$  for which  $\Psi_k(A) < \varphi_k(A)$ . In this section we further study these two approximation problems. We start with a geometrical characterization related to the function  $f(c, v)$  from (3.2).

THEOREM 5.1. *Let  $A \in \mathbb{R}^{n \times n}$  be a nonsingular matrix and let  $1 \leq k \leq d(A) - 1$ . The  $k$ th ideal and worst-case GMRES approximations are equal, i.e.,*

$$(5.1) \quad \max_{v \in S} \min_{c \in \mathbb{R}^k} f(c, v) = \min_{c \in \mathbb{R}^k} \max_{v \in S} f(c, v),$$

*if and only if  $f(c, v)$  has a saddle point in  $\mathbb{R}^k \times S$ .*

*Proof.* If  $f(c, v)$  has a saddle point in  $\mathbb{R}^k \times S$ , then there exist vectors  $\tilde{c} \in \mathbb{R}^k$  and  $\tilde{v} \in S$  such that

$$f(\tilde{c}, v) \leq f(\tilde{c}, \tilde{v}) \leq f(c, \tilde{v}) \quad \forall c \in \mathbb{R}^k, \forall v \in S.$$

The condition  $f(\tilde{c}, v) \leq f(\tilde{c}, \tilde{v})$  for all  $v \in S$  implies that  $\tilde{v}$  is a maximal right singular vector of the matrix  $p(A; \tilde{c})$ . If  $f(\tilde{c}, \tilde{v}) \leq f(c, \tilde{v})$  for all  $c \in \mathbb{R}^k$ , then  $p(z; \tilde{c})$  is the GMRES polynomial that corresponds to the initial vector  $\tilde{v}$ . In other words, if  $f(c, v)$  has a saddle point in  $\mathbb{R}^k \times S$ , then there exist a polynomial  $p(z; \tilde{c})$  and a unit norm vector  $\tilde{v}$  such that  $\tilde{v}$  is a maximal right singular vector of  $p(A; \tilde{c})$  and

$$p(A; \tilde{c})\tilde{v} \perp A\mathcal{K}_k(A, \tilde{v}).$$

Using [12, Lemma 2.4], the  $k$ th ideal and worst-case GMRES approximations are then equal.

On the other hand, if the condition (5.1) is satisfied, then  $f(c, v)$  has a saddle point in  $\mathbb{R}^k \times S$ .  $\square$

In other words, the  $k$ th ideal and worst-case GMRES approximations are equal if and only if the points  $(\tilde{c}, \tilde{v}) \in \mathbb{R}^k \times S$  that solve the worst-case GMRES problem are also the saddle points of  $f(c, v)$  in  $\mathbb{R}^k \times S$ .

We next extend the original construction of Toh [13] to obtain some further numerical examples in which  $\Psi_k(A) < \varphi_k(A)$ . Note that the Toh matrix (4.1) is not diagonalizable. In particular, for  $\omega = 1$  we have  $A = X\tilde{J}X^{-1}$ , where

$$\tilde{J} = \begin{bmatrix} 1 & 1 & & \\ & 1 & & \\ & & -1 & 1 \\ & & & -1 \end{bmatrix}, \quad X = \begin{bmatrix} \epsilon & \epsilon & \epsilon & -\epsilon \\ -2 & -1 & 0 & 1 \\ 0 & -2\epsilon & 0 & 2\epsilon \\ 0 & 4 & 0 & 0 \end{bmatrix}.$$

One can ask whether the phenomenon  $\Psi_k(A) < \varphi_k(A)$  can appear also for diagonalizable matrices. The answer is yes, since both  $\Psi_k(A)$  and  $\varphi_k(A)$  are continuous functions on the open set of nonsingular matrices; see [1, Theorem 2.5 and Theorem 2.6]. Hence one can slightly perturb the diagonal of the Toh matrix (4.1) in order to obtain a diagonalizable matrix  $\tilde{A}$  for which  $\Psi_k(\tilde{A}) < \varphi_k(\tilde{A})$ .

For  $\omega = 1$ , the Toh matrix is an upper bidiagonal matrix with the alternating diagonal entries 1 and  $-1$ , and the alternating superdiagonal entries  $\epsilon$  and  $\epsilon^{-1}$ . One can consider such a matrix for any  $n \geq 4$ , i.e.,

$$A = \begin{bmatrix} 1 & \epsilon & & & \\ & -1 & \epsilon^{-1} & & \\ & & 1 & \epsilon & \\ & & & \ddots & \ddots \\ & & & & \ddots & \epsilon^{\pm 1} \\ & & & & & \pm 1 \end{bmatrix} \in \mathbb{R}^{n \times n},$$

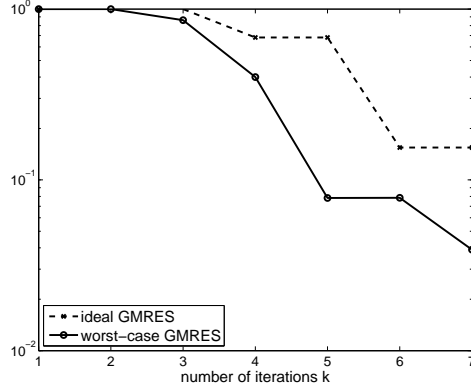


FIG. 5.1. *Ideal and worst-case GMRES can differ from step 3 up to the step  $2n - 1$ .*

and look at the values of  $\Psi_k(A)$  and  $\varphi_k(A)$ . If  $n$  is even, we found numerically that  $\Psi_k(A) = \varphi_k(A)$  for  $k \neq n-1$  and  $\Psi_{n-1}(A) < \varphi_{n-1}(A)$ . If  $n$  is odd, then our numerical experiments showed that  $\Psi_k(A) = \varphi_k(A)$  for  $k \neq n-2$  and  $\Psi_{n-2}(A) < \varphi_{n-2}(A)$ . Hence for all such matrices worst-case and ideal GMRES differ from each other for exactly one  $k$ .

Inspired by the Toh matrix, we define the  $n \times n$  matrices (for any  $n \geq 2$ )

$$J_{\lambda, \varepsilon} \equiv \begin{bmatrix} \lambda & \varepsilon & & \\ & \ddots & \ddots & \\ & & \ddots & \varepsilon \\ & & & \lambda \end{bmatrix}, \quad E_\varepsilon \equiv \begin{bmatrix} 0 & 0 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & 0 \\ \varepsilon^{-1} & 0 & \dots & 0 \end{bmatrix}$$

and use them to construct the matrix

$$A = \begin{bmatrix} J_{1, \varepsilon} & \omega E_\varepsilon \\ J_{-1, \varepsilon} & \end{bmatrix} \in \mathbb{R}^{2n \times 2n}, \quad \omega > 0.$$

One can numerically observe that here  $\Psi_k(A) < \varphi_k(A)$  for all steps  $k = 3, \dots, 2n-1$ . As an example, we plot in Fig. 5.1 the ideal and worst-case GMRES convergence curves for  $n = 4$ , i.e.,  $A$  is an  $8 \times 8$  matrix,  $\omega = 4$  and  $\varepsilon = 0.1$ . Varying the parameter  $\omega$  will influence the difference between worst-case and ideal GMRES in these examples.

## 6. Ideal and worst-case GMRES for complex vectors or polynomials.

We now ask whether the values of the max-min approximation (1.3) and the min-max approximation (1.4) for a matrix  $A \in \mathbb{R}^{n \times n}$  can change if we allow the maximization over complex vectors and/or the minimization over complex polynomials. The answer to this question will show that the two approximation problems indeed are of a different nature.

Let us define

$$\varphi_{k, \mathbb{K}, \mathbb{F}}(A) \equiv \min_{p \in \pi_{k, \mathbb{K}}} \max_{\substack{b \in \mathbb{F}^n \\ \|b\|=1}} \|p(A)b\|, \quad \Psi_{k, \mathbb{K}, \mathbb{F}}(A) \equiv \max_{\substack{b \in \mathbb{F}^n \\ \|b\|=1}} \min_{p \in \pi_{k, \mathbb{K}}} \|p(A)b\|,$$

where  $\mathbb{K}$  and  $\mathbb{F}$  are either the real or the complex numbers. Hence, the previously used  $\varphi_k(A)$ ,  $\Psi_k(A)$ , and  $\pi_k$  are now denoted by  $\varphi_{k, \mathbb{R}, \mathbb{R}}(A)$  and  $\Psi_{k, \mathbb{R}, \mathbb{R}}(A)$ , and  $\pi_{k, \mathbb{R}}$ , respectively. We first analyze the case of  $\varphi_{k, \mathbb{K}, \mathbb{F}}(A)$ .

THEOREM 6.1. For a nonsingular matrix  $A \in \mathbb{R}^{n \times n}$  and  $1 \leq k \leq d(A) - 1$ ,

$$\varphi_{k,\mathbb{R},\mathbb{R}}(A) = \varphi_{k,\mathbb{C},\mathbb{R}}(A) = \varphi_{k,\mathbb{R},\mathbb{C}}(A) = \varphi_{k,\mathbb{C},\mathbb{C}}(A).$$

*Proof.* Since

$$\max_{\substack{b \in \mathbb{R}^n \\ \|b\|=1}} \|Bv\| = \|B\| = \max_{\substack{b \in \mathbb{C}^n \\ \|b\|=1}} \|Bv\|$$

holds for any real matrix  $B \in \mathbb{R}^{n \times n}$ , we have  $\varphi_{k,\mathbb{R},\mathbb{R}}(A) = \varphi_{k,\mathbb{R},\mathbb{C}}(A)$ .

Next, from  $\mathbb{R} \subset \mathbb{C}$  we get immediately  $\varphi_{k,\mathbb{C},\mathbb{R}}(A) \leq \varphi_{k,\mathbb{R},\mathbb{R}}(A)$ . On the other hand, writing  $p \in \pi_{k,\mathbb{C}}$  in the form  $p = p_r + \mathbf{i}p_i$ , where  $p_r \in \pi_{k,\mathbb{R}}$  and  $p_i$  is a real polynomial of degree at most  $k$  such that  $p_i(0) = 0$ , we get

$$\begin{aligned} \varphi_{k,\mathbb{C},\mathbb{R}}^2(A) &= \min_{p \in \pi_{k,\mathbb{C}}} \max_{\substack{b \in \mathbb{R}^n \\ \|b\|=1}} \|p(A)b\|^2 = \min_{p \in \pi_{k,\mathbb{C}}} \max_{\substack{b \in \mathbb{R}^n \\ \|b\|=1}} (\|p_r(A)b\|^2 + \|p_i(A)b\|^2) \\ &\geq \min_{p_r \in \pi_{k,\mathbb{R}}} \max_{\substack{b \in \mathbb{R}^n \\ \|b\|=1}} \|p_r(A)b\|^2 = \varphi_{k,\mathbb{R},\mathbb{R}}^2(A), \end{aligned}$$

so that  $\varphi_{k,\mathbb{C},\mathbb{R}}(A) = \varphi_{k,\mathbb{R},\mathbb{R}}(A)$ . Finally, from [5, Theorem 3.1] we obtain  $\varphi_{k,\mathbb{R},\mathbb{R}}(A) = \varphi_{k,\mathbb{C},\mathbb{C}}(A)$ .  $\square$

Since the value of  $\varphi_{k,\mathbb{K},\mathbb{F}}(A)$  does not change when choosing for  $\mathbb{K}$  and  $\mathbb{F}$  real or complex numbers, we will again use the simple notation  $\varphi_k(A)$  in the following text. The situation for the quantities corresponding to the worst-case GMRES approximation is more complicated. Our proof of this fact uses the following lemma.

LEMMA 6.2. If  $A = A(\omega, \varepsilon)$  is the Toh matrix defined in (4.1) and

$$(6.1) \quad B \equiv \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix},$$

then  $\Psi_{3,\mathbb{R},\mathbb{R}}(B) = \varphi_3(A)$ .

*Proof.* Using the structure of  $B$  it is easy to see that  $\Psi_{k,\mathbb{R},\mathbb{R}}(B) \leq \varphi_k(A)$  for any  $k$ . To prove the equality, it suffices to find a real unit norm vector  $w$  with

$$(6.2) \quad \min_{p \in \pi_{3,\mathbb{R}}} \|p(B)w\| = \varphi_3(A) = \min_{p \in \pi_{3,\mathbb{R}}} \|p(A)\|.$$

The solution  $p_*$  of the ideal GMRES problem on the right hand side of (6.2) is given by (4.3). Toh showed in [13, p. 32] that  $p_*(A)$  has a twofold maximal singular value  $\sigma$ , and that the corresponding right and left singular vectors are given (up to a normalization) by

$$[v_1, v_2] = \begin{bmatrix} 0 & c \\ c & 0 \\ 0 & -2 \\ -2 & 0 \end{bmatrix}, \quad [u_1, u_2] = \begin{bmatrix} 0 & 2 \\ 2 & 0 \\ 0 & -c \\ -c & 0 \end{bmatrix},$$

i.e.,  $\sigma u_1 = p_*(A)v_1$  and  $\sigma u_2 = p_*(A)v_2$ , where  $\sigma = \|p_*(A)\|$ .

Let us define

$$w \equiv \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} / \left\| \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\|, \quad q(z) \equiv p_*(z).$$

Using

$$q(B) \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \sigma \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad \text{and} \quad \left\| \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\| = \left\| \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \right\|,$$

we see that  $\|q(B)w\| = \sigma$ . To prove (6.2) it is sufficient to show that  $q$  is the third GMRES polynomial for  $B$  and  $w$ , i.e., that  $q$  satisfies  $q(B)w \perp B^j w$  for  $j = 1, 2, 3$ , or, equivalently,

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}^T \begin{bmatrix} A^j & 0 \\ 0 & A^j \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = u_1^T A^j v_1 + u_2^T A^j v_2 = 0, \quad j = 1, 2, 3.$$

Using linear algebra calculations we get  $u_1^T A v_1 = -4c = -u_2^T A v_2$ , and

$$0 = u_1^T A^2 v_1 = u_2^T A^2 v_2 = u_1^T A^3 v_1 = u_2^T A^3 v_2.$$

Therefore, we have found a unit norm initial vector  $w$  and the corresponding third GMRES polynomial  $q$  such that  $\|q(B)w\| = \varphi_3(A)$ .  $\square$

We next analyze the quantities  $\Psi_{k,\mathbb{K},\mathbb{F}}(A)$ .

**THEOREM 6.3.** *For a nonsingular matrix  $A \in \mathbb{R}^{n \times n}$  and  $1 \leq k \leq d(A) - 1$ ,*

$$\Psi_{k,\mathbb{R},\mathbb{R}}(A) = \Psi_{k,\mathbb{C},\mathbb{R}}(A) \leq \Psi_{k,\mathbb{C},\mathbb{C}}(A) \leq \Psi_{k,\mathbb{R},\mathbb{C}}(A),$$

where both inequalities can be strict.

*Proof.* For a real initial vector  $b$ , the corresponding GMRES polynomial is uniquely determined and real. This implies  $\Psi_{k,\mathbb{C},\mathbb{R}}(A) = \Psi_{k,\mathbb{R},\mathbb{R}}(A)$ . Next, from [5, Theorem 3.1] it follows that  $\Psi_{k,\mathbb{R},\mathbb{R}}(A) \leq \Psi_{k,\mathbb{C},\mathbb{C}}(A)$ . Finally, using  $\mathbb{R} \subset \mathbb{C}$  we get  $\Psi_{k,\mathbb{C},\mathbb{C}}(A) \leq \Psi_{k,\mathbb{R},\mathbb{C}}(A)$ .

It remains to show that the inequalities can be strict. For the first inequality, as shown in [16, Section 4], there exist real matrices  $A$  and certain complex (unit norm) initial vectors  $b$  for which  $\min_{p \in \pi_{k,\mathbb{C}}} \|p(A)b\| = 1$  for  $k = 1, \dots, n-1$  (complete stagnation), while such complete stagnation does not occur for any real (unit norm) initial vector. Therefore, there are matrices for which  $\Psi_{k,\mathbb{C},\mathbb{R}}(A) < \Psi_{k,\mathbb{C},\mathbb{C}}(A)$ .

To show that the second inequality can be strict, we note that for any  $A \in \mathbb{R}^{n \times n}$ , the corresponding matrix  $B \in \mathbb{R}^{2n \times 2n}$  of the form (6.1), and  $1 \leq k \leq d(A) - 1$ ,

$$\begin{aligned} \Psi_{k,\mathbb{R},\mathbb{C}}^2(A) &= \max_{\substack{b \in \mathbb{C}^n \\ \|b\|=1}} \min_{p \in \pi_{k,\mathbb{R}}} \|p(A)b\|^2 = \max_{\substack{u, v \in \mathbb{R}^n \\ \|u\|^2 + \|v\|^2 = 1}} \min_{p \in \pi_{k,\mathbb{R}}} \|p(A)(u + \mathbf{i}v)\|^2 \\ &= \max_{\substack{u, v \in \mathbb{R}^n \\ \|u\|^2 + \|v\|^2 = 1}} \min_{p \in \pi_{k,\mathbb{R}}} (\|p(A)u\|^2 + \|p(A)v\|^2) \\ (6.3) \quad &= \max_{\substack{v \in \mathbb{R}^{2n} \\ \|v\|=1}} \min_{p \in \pi_{k,\mathbb{R}}} \|p(B)v\|^2 = \Psi_{k,\mathbb{R},\mathbb{R}}^2(B). \end{aligned}$$

Now let  $A$  be the Toh matrix (4.1) and  $k = 3$ . Toh showed in [13, Theorem 2.2] that for any unit norm  $b \in \mathbb{C}^4$  and the corresponding third GMRES polynomial  $p_b \in \pi_{3,\mathbb{C}}$ ,

$$\|p_b(A)b\| < \varphi_3(A).$$

Hence  $\Psi_{3,\mathbb{C},\mathbb{C}}(A) < \varphi_3(A)$ . Lemma 6.2 and equation (6.3) imply  $\varphi_3(A) = \Psi_{3,\mathbb{R},\mathbb{C}}(A)$ , which completes the proof of the strict inequality.  $\square$



Our proof concerning the strictness of the first inequality in the previous theorem relied on a numerical example given in [16, Section 4]. We will now give an alternative construction based on the non-uniqueness of the worst-case GMRES polynomial, which will lead to an example with

$$\Psi_{k,\mathbb{R},\mathbb{R}}(A) < \Psi_{k,\mathbb{R},\mathbb{C}}(A).$$

Suppose that  $A$  is a real matrix for which in a certain step  $k$  two *different* worst-case polynomials  $p_b \in \pi_{k,\mathbb{R}}$  and  $p_c \in \pi_{k,\mathbb{R}}$  with corresponding real unit norm initial vectors  $b$  and  $c$  exist, so that

$$\Psi_{k,\mathbb{R},\mathbb{R}}(A) = \|p_b(A)b\| = \|p_c(A)c\|.$$

Note that since  $p_b$  and  $p_c$  are the uniquely determined GMRES polynomials that solve the problem (1.1) for the corresponding real initial vectors, it holds that

$$(6.4) \quad \|p_b(A)b\| < \|p(A)b\|, \quad \|p_c(A)c\| < \|p(A)c\|$$

for any polynomial  $p \in \pi_{k,\mathbb{C}} \setminus \{p_b, p_c\}$ .

Writing any complex vector  $w \in \mathbb{C}^n$  in the form  $w = (\cos \theta) u + \mathbf{i}(\sin \theta) v$ , with  $u, v \in \mathbb{R}^n$ ,  $\|u\| = \|v\| = 1$ , we get

$$\begin{aligned} \Psi_{k,\mathbb{R},\mathbb{C}}^2(A) &= \max_{\substack{w \in \mathbb{C}^n \\ \|w\|=1}} \min_{p \in \pi_{k,\mathbb{R}}} \|p(A)b\|^2 \\ &= \max_{\substack{\theta \in \mathbb{R}, u, v \in \mathbb{R}^n \\ \|u\|=\|v\|=1}} \min_{p \in \pi_{k,\mathbb{R}}} (\cos^2 \theta \|p(A)u\|^2 + \sin^2 \theta \|p(A)v\|^2) \\ &\geq \max_{\theta \in \mathbb{R}} \min_{p \in \pi_{k,\mathbb{R}}} (\cos^2 \theta \|p(A)b\|^2 + \sin^2 \theta \|p(A)c\|^2) \\ &> (\cos^2 \theta) \Psi_{k,\mathbb{R},\mathbb{R}}^2(A) + (\sin^2 \theta) \Psi_{k,\mathbb{R},\mathbb{R}}^2(A) = \Psi_{k,\mathbb{R},\mathbb{R}}^2(A), \end{aligned}$$

where the strict inequality follows from (6.4) and from the fact that  $\|p(A)b\|^2$  and  $\|p(A)c\|^2$  do not attain their minima for the same polynomial.

To demonstrate the strict inequality  $\Psi_{k,\mathbb{R},\mathbb{R}}(A) < \Psi_{k,\mathbb{R},\mathbb{C}}(A)$  numerically we use the Toh matrix (4.1) with  $\varepsilon = 0.1$  and  $\omega = 1$ , and  $k = 3$ . Let  $b$  and  $c$  be the corresponding two different worst-case initial vectors introduced in Section 4. We vary  $\theta$  from 0 to  $\pi$  and compute the quantities

$$(6.5) \quad \min_{p \in \pi_{3,\mathbb{R}}} (\cos^2 \theta \|p(A)b\|^2 + \sin^2 \theta \|p(A)c\|^2) = \min_{p \in \pi_{3,\mathbb{R}}} \|p(B)g_\theta\|^2,$$

where

$$B = \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix} \quad \text{and} \quad g_\theta = \begin{bmatrix} (\cos \theta)b \\ (\sin \theta)c \end{bmatrix}.$$

In Fig. 6.1 we can see clearly, that for  $\theta \notin \{0, \pi/2, \pi\}$  the value of (6.5) is strictly larger than  $\Psi_3(A) = 0.4579$ .

**7. Concluding remarks.** We have studied the worst-case GMRES approximation problem, which for each (nonsingular) matrix  $A$  and iteration step  $k \leq d(A)$  represents the best possible attainable upper bound on the actual GMRES residual norm for a linear algebraic system with  $A$  at step  $k$ . We have derived several theoretical properties of the worst-case GMRES problem, and we have studied its relation to the ideal GMRES approximation problem.

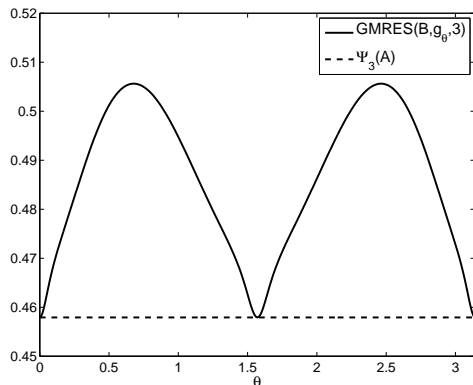


FIG. 6.1. The GMRES residual norm for a varying complex right hand side.

In this paper we did not consider quantitative estimation of the worst-case GMRES value  $\Psi_k(A)$ , and we did not study how this value depends on properties of  $A$ . This is an important problem of great practical interest, which is largely open. For more details and a survey of the current state-of-the-art we refer to [8, Section 5.7].

#### REFERENCES

- [1] V. FABER, W. JOUBERT, E. KNILL, AND T. MANTEUFFEL, *Minimal residual method stronger than polynomial preconditioning*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 707–729.
- [2] A. GREENBAUM, *Iterative Methods for Solving Linear Systems*, vol. 17 of Frontiers in Applied Mathematics, SIAM, Philadelphia, PA, 1997.
- [3] A. GREENBAUM AND L. GURVITS, *Max-min properties of matrix factor norms*, SIAM J. Sci. Comput., 15 (1994), pp. 348–358.
- [4] ANNE GREENBAUM AND LLOYD N. TREFETHEN, *GMRES/CR and Arnoldi/Lanczos as matrix approximation problems*, SIAM J. Sci. Comput., 15 (1994), pp. 359–368.
- [5] WAYNE JOUBERT, *On the convergence behavior of the restarted GMRES algorithm for solving nonsymmetric linear systems*, Numer. Linear Algebra Appl., 1 (1994), pp. 427–447.
- [6] ———, *A robust GMRES-based adaptive polynomial preconditioning algorithm for nonsymmetric linear systems*, SIAM J. Sci. Comput., 15 (1994), pp. 427–439.
- [7] PETER D. LAX, *Linear Algebra and its Applications*, Pure and Applied Mathematics (Hoboken), Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second ed., 2007.
- [8] J. LIESEN AND Z. STRAKOŠ, *Krylov Subspace Methods. Principles and Analysis*, Oxford University Press, Oxford, 2013.
- [9] JÖRG LIESEN AND PETR TICHÝ, *On best approximations of polynomials in matrices in the matrix 2-norm*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 853–863.
- [10] YOUSEF SAAD, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, PA, second ed., 2003.
- [11] YOUSEF SAAD AND MARTIN H. SCHULTZ, *GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.
- [12] PETR TICHÝ, JÖRG LIESEN, AND VANCE FABER, *On worst-case GMRES, ideal GMRES, and the polynomial numerical, hull of a Jordan block*, Electron. Trans. Numer. Anal., 26 (2007), pp. 453–473.
- [13] KIM-CHUAN TOH, *GMRES vs. ideal GMRES*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 30–36.
- [14] KIM-CHUAN TOH AND LLOYD N. TREFETHEN, *The Chebyshev polynomials of a matrix*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 400–419.
- [15] ILYA ZAVORIN, *Spectral factorization of the Krylov matrix and convergence of GMRES*, Tech. Report CS-TR-4309, Computer Science Department, University of Maryland, 2001.
- [16] ILYA ZAVORIN, DIANNE P. O’LEARY, AND HOWARD ELMAN, *Complete stagnation of GMRES*, Linear Algebra Appl., 367 (2003), pp. 165–183.